



Universidad Nacional de La Plata



Determinantes de la deserción y graduación universitaria: Una aplicación utilizando modelos de duración¹

Paula Inés Giovagnoli²

Documento de Trabajo Nro. 37

Marzo 2002

¹ Tesis de la Maestría en Economía de la UNLP dirigida por el Dr. Alberto Porto.

² paulagiovagnoli@hotmail.com

RESUMEN²

Este trabajo investiga el problema de la deserción y graduación de los estudiantes universitarios. Se utilizan modelos de riesgo proporcional no paramétricos para estimar los efectos cualitativos y cuantitativos de factores personales y características socioeconómica de los alumnos sobre las probabilidades condicionales de deserción y graduación. Se emplean datos de una cohorte de estudiantes de la carrera de Contador Público de la Universidad Nacional de Rosario, Argentina. Los hallazgos son útiles para evaluar y diseñar políticas públicas en el sector de educación.

Palabras claves: estudiantes universitarios - deserción y graduación

ABSTRACT

This study examines the issue of student departure from a public university. Non-parametric proportional hazard models are used to estimate the quantity and quality effects of the students' personal and socioeconomic characteristics on the probability of their dropping out or graduating. Data include a cohort of students who started studying accounting at National University of Rosario, Argentina, in 1991. The results are useful to evaluate and design public policies in the educational sector.

Keywords: university students – drop out and graduation

JEL: I-121, C-41

² Quiero agradecer al Profesor Alberto Porto por su permanente guía y apoyo en este trabajo. Especial mención para Norman K. Thurston, que me ha guiado especialmente en los aspectos econométricos y ha permitido enriquecerme a partir de valiosas discusiones. Agradezco a S. Jenkins por contestar mis dudas sobre el programa Stata. Agradezco a A. Castaña, Decana de la Facultad de Ciencias Económicas y Estadísticas de la Universidad Nacional de Rosario por brindarme los datos, al Ing. Aldo Gimbatti, Vicerrector de la Universidad Nacional de Rosario y Ana Inés Navarro, a C. Guarnieri y al Personal del Departamento de Alumnado; a M. Quaglino, W. Sosa Escudero y a R. Alemany Leira por el aporte bibliográfico.

Agradezco también los comentarios y sugerencias de los participantes del seminario de tesis.

ÍNDICE

I. INTRODUCCIÓN	1
II. ESTUDIOS PREVIOS	3
III. MODELOS DE DURACIÓN	6
1. Consideraciones metodológicas generales	6
2. Modelización	8
IV. DATOS	14
V. RESULTADOS EMPÍRICOS	17
1. Factores relacionados con el riesgo de deserción	17
2 Factores relacionados con la graduación	21
VI CONCLUSIONES	24
ANEXOS	26
A. Descripción de las variables	26
B. Reordenamiento de la base de datos	28
BIBLIOGRAFÍA	30
NOTAS	33

I. INTRODUCCIÓN

Las universidades públicas en la Argentina han implementado un sistema irrestricto de ingreso desde su creación. Exceptuando algunas etapas como por ejemplo entre los años 1977 y 1982, época en las que se aplicó un sistema de selección-admisión, en la mayoría de los demás años la universidad nacional ha gozado de un sistema no sólo de ingreso irrestricto sino también gratuito.

En parte esto ha generado un aumento considerable de la demanda en educación superior a lo largo de los años, especialmente a partir de la reinstauración de la democracia. Dicho aumento no ha estado acompañado por incrementos similares en la cantidad de graduados. Tomando como ejemplo la Universidad Nacional de Rosario (UNR), en 1973 la cantidad de ingresantes era aproximadamente 21.000 en tanto que los egresados apenas superaban los 2.300. En 1999 la cantidad de inscriptos alcanzaba ya los 67.000, mientras que los graduados se mantenían en alrededor de los 2500. Cifras relativas similares se observan para las demás Universidades Nacionales del país¹.

Esto refleja la existencia de dos fenómenos claros. Por un lado, el grupo de graduados es reducido en relación con la cantidad de ingresantes. Por el otro, es evidente que gran parte del grupo que inicia sus estudios abandona la carrera. Además, como bien se muestra en varias investigaciones (Pagura et al., 2000; Porto y Di Gresia, 2001) dentro del grupo que llega a graduarse, existe un importante porcentaje de alumnos que prolonga su permanencia en la universidad más allá de los tiempos establecidos por los planes de estudios.

Algunos autores, por ejemplo Manski (1989) argumentan que el abandono no es necesariamente malo ya que “el sólo hecho de haber iniciado la universidad permite a la persona tener información sobre su adaptación al entrenamiento universitario”. Adicionalmente, podría pensarse que el iniciar una carrera agrega conocimientos con valor económico, al menos para los años que el alumno asistió a la universidad.

Sin embargo, más allá de la discusión planteada en el párrafo anterior, interesa saber que tiempo le lleva al individuo tomar la decisión de desertar o graduarse y, sobre todo, cuáles son los factores que influyen en tal decisión. Resulta de particular importancia determinar si la probabilidad de abandonar los estudios o graduarse es la misma a lo largo de la trayectoria académica de un alumno.

En nuestro país son pocas las investigaciones que intentan dar respuestas a estos problemas. Sin embargo, el análisis y resolución de este tipo de cuestiones es importante tanto para las mismas universidades como para el Estado, que es quien provee los recursos monetarios necesarios para financiar la educación superior. Poder identificar los años de mayores riesgos de abandono posibilita diseñar políticas que prevengan la deserción y, a su vez, lograr una administración más eficiente de los recursos escasos. Y aun más interesante es conocer, por ejemplo, si el riesgo de desertar es significativamente mayor para un ingresante cuyos padres tienen bajos niveles educativos y ocupaciones relativamente poco remuneradas versus estudiantes que se encuentran en mejores condiciones socioeconómicas. Medir esta diferencia, si es que existe, adiciona una información útil a la problemática universitaria antes descripta.

Esta investigación plantea dos objetivos principales. El primero es investigar *cundo* es más probable que un alumno abandone o se gradúe en la universidad. El segundo objetivo es identificar *cuáles* son las características socioeconómicas y los factores personales que más se relacionan con la duración y riesgo de cada evento.

Para lograr estos objetivos, se aplica el método estadístico denominado "análisis de transición", que permite estudiar el problema de la deserción y la graduación a partir de información disponible sobre la duración de cada evento- que en la mayoría de los casos es incompleta, ya que el período de observación no es infinito.

Se realiza una aplicación empírica utilizando los datos de la Facultad de Ciencias Económicas y Estadística de la UNR. Se estudia la cohorte compuesta por estudiantes que ingresaron a la carrera de Contador Público en 1991. Se dispone de información sobre características personales de los alumnos, condiciones socioeconómicas de las familias y el rendimiento académico desde el ingreso hasta su graduación, deserción o continuación de los estudios.

Uno de los aportes de este trabajo con respecto a otros estudios es que la mayoría de las investigaciones ya realizadas, como se expondrá en el apartado II, examinan si el evento (la deserción o la graduación) ocurre en un determinado momento del tiempo. En cambio, la técnica aquí aplicada permite ver al evento como un proceso en el tiempo, evaluando quiénes son los estudiantes que tienen más propensión (mayor probabilidad) a abandonar o recibirse *teniendo en cuenta el tiempo que llevan estudiando*. Se modela además el evento de finalización de los estudios distinguiendo entre culminar por causa de deserción o bien por obtención del título universitario (modelo de "múltiples causas"). En la sección III se analiza con detalle la metodología empleada y se justifica su elección. La sección IV describe los datos disponibles y explica el porqué se trabaja con una cohorte. La sección V analiza los resultados obtenidos. La última sección esta destinada a enunciar las conclusiones del trabajo y direccionar futuras investigaciones.

II. ESTUDIOS PREVIOS

Gran parte de la literatura referida a la permanencia² de los estudiantes en la universidad se desarrolla sobre la base de dos principales teorías sociológicas: el modelo de integración del estudiante (Student Integration Model) Spady (1970), Tinto (1975) y el modelo de desgaste del estudiante (Student Attrition Model) Bean (1980). El primero de los modelos explica que, dado lo demás constante, un mayor grado de integración del estudiante en el ambiente académico y social contribuye a un mayor grado de compromiso institucional y esto afecta directamente a la decisión del alumno de permanecer o desertar. El segundo modelo atribuye mayor importancia relativa a los factores externos a la institución. Cabrera, et al. (1993) consideran que si bien estos modelos parecieran ser opuestos, en realidad existe complementariedad entre ellos y a su vez superposición en algunos aspectos. Es así como proponen un modelo integrado que enfatiza los procesos sociológicos y psicológicos del comportamiento de la persistencia del alumno en la universidad.

Singer y Willet (1991) realizan una importante revisión sistemática de varios trabajos empíricos basados en estos modelos utilizando diferentes metodologías.

Las investigaciones muestran que los alumnos que desertan, comparados con aquellos que permanecen en la universidad, tienden a tener menores notas académicas en promedio y padres con menor educación e inferiores ingresos. También se ha estimado que los estudiantes con un mayor nivel de interacción con el claustro de profesores y con otros estudiantes tienen menor probabilidad de abandonar, pero es muy posible que tal interacción sea una función de muchos de los factores que influyen en la probabilidad de desertar, Clotfelter C. T. et al. (1991). Se ha presentado que la deserción es mayor en el primer año de la carrera universitaria Robinson, R. (1990).

La identificación de los grupos, desertores y graduados, y el cálculo de la probabilidad de pertenecer a uno u otro conjunto, dadas ciertas características, permiten diseñar políticas de permanencia, maximizando así el uso de los recursos disponibles en las universidades y minimizando los costos sociales³

Sin embargo, los trabajos anteriores abordan el problema de la deserción o graduación bajo un marco estático. Es decir, sólo investigan si ocurre o no el suceso y de qué factores depende este hecho pero ignoran *cundo* ocurre. En otras palabras, estos estudios no permiten captar la evolución del evento a lo largo del tiempo.

Para incluir de alguna manera una dimensión más dinámica al análisis, algunos autores han trabajado con modelos dicotómicos sucesivos (por ejemplo un logit en distintos momentos del tiempo) para comparar la probabilidad de abandonar (recibirse) en cada período Schlechty y Vance (1981) o bien analizar los factores que explican el suceso empleando una regresión logística para los años académicos sucesivos Alemany et al. (1990). Estos trabajos concluyen que la probabilidad de finalizar (o abandonar) no es constante a lo largo de la vida académica de un estudiante.

Adelman (1999) utilizando datos longitudinales obtenidos del Centro Nacional de Estadísticas Educativas de Estados Unidos, analiza a una cohorte de estudiantes ingresantes a una Universidad para identificar los factores que contribuyen a que el alumno se gradúe. Incorpora una gran cantidad de variables explicativas relacionadas con el desempeño de cada estudiante, los resultados de los exámenes de ingresos y la información disponible sobre sus aptitudes. A partir de un modelo de regresión logística encuentra que los recursos académicos y el patrón de asistencia son los factores que

más explican la obtención del título universitario. Nuevamente, el trabajo no considera como los factores podrían influenciar en el *tiempo* necesario hasta lograr la graduación.

DesJardins et al. (2001) investigan este último aspecto. Estos autores tratan la salida de un estudiante de la universidad como un “proceso” que se verifica a lo largo del tiempo y enfatizan en que “los diseños estáticos no incorporan adecuadamente la dimensión temporal de este proceso”. Aplicando una metodología alternativa denominada análisis de transición o duración⁴ a la misma base de datos utilizada por Adelman, concluyen que el GPA (Grade Point Average) es el mejor predictor de la obtención del título universitario. Específicamente, encuentran que un aumento en un punto del GPA aumenta más del doble la probabilidad que tiene un estudiante de recibirse. Sin embargo, este efecto cae a medida que pasa el tiempo.

Montoya Díaz (1999), adopta una concepción utilitarista para modelizar la dinámica del comportamiento⁵ de los estudiantes de una universidad de Brasil. Aplicando un modelo de duración concluye, entre otras cosas, que cuanto mayor es el ingreso (familiar) mayor es el tiempo que el alumno estará en la facultad antes de completar sus estudios y menor es el tiempo antes de abandonar. Esto muestra que entre los estudiantes que prolongan su permanencia en la universidad, aquellos que cuentan con menores ingresos al iniciar sus estudios tienen mayores probabilidades de desertar, mientras que los alumnos con mayores ingresos tienen mayores probabilidades de completar.

La mayoría de los estudios consultados están realizados para universidades de Estados Unidos (que no tienen un sistema libre y gratuito) y sólo unos pocos centran su interés en Universidades latinoamericanas.

En la Argentina, el problema de fracaso y deserción universitaria se plantea desde los inicios de la Universidad Pública. En 1968, se enfatizaba sobre la importancia de investigar “la composición social de los estudiantes y graduados y el problema de la deserción *hoy día* tan agudo” Araoz (1968). Sin embargo, nuevamente se evidencian muy pocos estudios realizados en esta área.

Porto et al. (2000) analizaron el rendimiento de los estudiantes universitarios y sus determinantes⁶ empleando diferentes indicadores como por ejemplo la razón entre la cantidad de materias rendidas y los años desde el ingreso; este mismo cociente ponderado por el promedio de materias; y distancia entre el rendimiento real y el rendimiento teórico. Encuentran, entre otras cosas, que la retención de alumnos con padres de menor nivel de educación es muy baja.

La Dirección de Estadística Universitaria de la Universidad Nacional de Rosario (UNR) realizó una investigación de corte transversal para el año 1990 Perfil Social (1990) aplicando técnicas de análisis factorial a la población total de alumnos de dicha Universidad. Se obtuvieron resultados para cada Facultad y luego se ajustaron las técnicas para llegar a resultados generales. Construyendo una topología de los estudiantes encontraron tres grupos con mayor riesgo de abandono. Estos grupos son: alumnos que trabajan 36 horas semanales y están casados; alumnos que trabajan y sus padres (obreros o comerciantes) tienen sólo nivel primario; estudiantes que viven en Rosario trabajan más de 20 horas con padres que han alcanzado educación primaria y son empleados o pequeños comerciantes. Contrariamente, el grupo que tiene menos riesgo de abandono está caracterizado por alumnos que no trabajan, provienen de otras zonas y sus padres tienen nivel secundario y se dedican a la rama agropecuaria.

Pagura et al. (2000) evalúan el tiempo demandado por un grupo de estudiantes de la Facultad de Ciencias Económicas y Estadísticas de la UNR para concluir los dos primeros años de una carrera y relacionan esta duración con el género, nivel de educación de los padres, condiciones de ingreso y otras variables socio-culturales. Estudian también los factores que influyen sobre la probabilidad de cumplir dicha etapa. Concluyen que las características que favorecen el rendimiento son: la aprobación de los exámenes de ingresos, el nivel educacional de los padres y el comienzo de la carrera universitaria inmediatamente después de haber terminado el nivel secundario.

Cerioni et al. (1999) basándose en una encuesta realizada a un grupo de desertores de la Universidad Nacional del Sur estimaron un modelo logit para identificar las variables explicativas de la probabilidad de un desertor de pertenecer a los dos quintiles de ingresos más altos. También construyeron indicadores para determinar las posibilidades que tiene un alumno de graduarse. Un resultado es que las mayores oportunidades de graduación se ubican entre aquellos estudiantes cuyas familias pertenecen a los estratos socioeconómicos elevados. Concluyen que “si bien las oportunidades de ingresar a la universidad pública son irrestrictas, el peso relativo sustancialmente mayor de estudiantes de niveles de ingresos medios y altos frente al correspondiente a los más bajos pone de manifiesto que los mecanismos redistributivos por medio del gasto social en educación superior no logren de manera óptima transferir ingresos sólo hacia los grupos económicamente más pobres.”

III. MODELOS DE DURACIÓN

1. Consideraciones metodológicas generales

El objetivo general del este trabajo es investigar *cuándo* es probable que ocurra un evento determinado (desertar o graduarse). Específicamente, interesa el tiempo que lleva tomar esta decisión y cuáles son las variables que más influyen en el cumplimiento del evento.

La técnica que permite describir el comportamiento de datos que corresponden al tiempo o duración desde un origen bien definido hasta la ocurrencia de algún evento o punto final se denomina “análisis de duración”; (Cox, 1972; Cox y Oakes, 1984; Lancaster, 1979; Kalbfleisch y Prentice, 1980; Klein. y Moeschberger, 1997) y es la aplicada en esta investigación⁷.

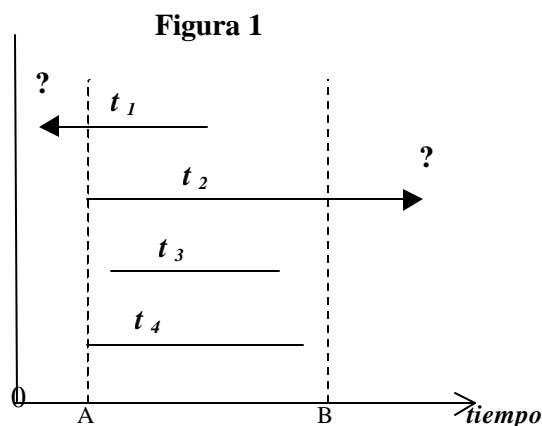
Esta técnica posee varias **ventajas** con respecto a técnicas clásicas como la estimación de modelos logit “clásicos”, de regresión o análisis discriminante. Estas últimas son de naturaleza estática, mientras que el análisis de duración capta la temporalidad y la variación de las circunstancias a lo largo del tiempo, siendo un enfoque más dinámico. El concepto central de un modelo de duración no es la probabilidad de que un evento ocurra (ej. probabilidad de que un individuo abandone sus estudios el tercer año) , sino mas bien la probabilidad *condicional* de que esto ocurra (probabilidad de que el estudiante abandone en el tercer año, *dado que ha estado estudiando por dos años*).

Permite además incluir en el modelo factores explicativos constantes y *variables* en el tiempo. En las técnicas clásicas, en cambio, al querer introducir factores cambiantes en el tiempo surgen problemas de colinealidad y autocorrelación, haciéndose necesario la corrección de estos inconvenientes Allison (1982).

Adicionalmente, en el análisis de duración se considera el tratamiento de “**ciertas cuestiones**” o problemas que usualmente surgen al tratar de modelar la variable de interés T que es una variable aleatoria no negativa que representa la duración hasta que ocurra un determinado evento. En principio, pueden enumerarse cuatro cuestiones importantes que incorpora esta metodología:

A) Censura

El problema principal que hace necesario el uso de modelos de duración es la existencia de censura en los datos. La censura ocurre cuando el resultado o evento de interés (por ejemplo graduarse o desertar) no se observa para todos los individuos dentro del período en que se realiza la recolección de los datos. Por ejemplo, si se está analizando el tiempo que le lleva a los individuos tomar la decisión de graduarse, se puede conocer con exactitud cuándo los individuos iniciaron la carrera (inscripción a la universidad). Sin embargo, si se supone que la medición de la duración de la carrera se realiza diez años después, en ese momento es posible que existan alumnos que aún estén cursando sus estudios por lo que la verdadera duración no es observada para ese grupo de estudiantes. En este caso se dice que el dato se encuentra *censurado por derecha*. En la **Figura 1** dicha situación viene representada por t_2 . En esta figura, A y B delimitan el inicio y finalización del período de observación, respectivamente.



El caso t_1 se da cuando, en el momento de iniciar la investigación, se observa la situación final del individuo pero no puede saberse cuando inició, y por lo tanto, no se tiene completa información de la duración del evento. Este caso se denomina *censura por izquierda*.

Para los individuos t_3 y t_4 no hay censura ya que se tiene información completa dentro del período de observación.

Cuando la censura no está relacionada con la ocurrencia del evento, se denomina “censura independiente”⁸. Si la censura no fuese independiente, las observaciones no censuradas diferirían “sistemáticamente” de las observaciones censuradas.

B) Tratamiento continuo versus discreto

La variable aleatoria T puede definirse como variable continua o discreta. Dicha definición lleva a diferentes tratamientos de estimación.

Si se considera que el tiempo es absolutamente continuo, T puede tomar valores positivos desde 0 a infinito. En este caso, se espera que no exista gran cantidad de observaciones con la misma duración.

Si en cambio, el tiempo se considera discreto, T toma valores enteros $0 < t_1 < t_2 < t_3 < \dots$. La consideración de “tiempos discretos” puede aplicarse a dos casos diferentes :

- ✓ Cuando la medida del tiempo es estrictamente discreta. Es decir, el registro de los datos de duración se hace en tiempos discretos porque el evento sólo puede ocurrir regularmente. Por ejemplo, las elecciones presidenciales son regulares cada determinado período de tiempo, en una fecha determinada.
- ✓ Cuando se trabaja con tiempos continuos pero registrados en forma de intervalos. En estos casos, el evento podría ocurrir en cualquier momento del tiempo, pero en el momento de recolección de los datos sólo se sabe si el evento ocurrió en determinado intervalo de tiempo. Los datos con los que se disponen en este trabajo son un ejemplo de este tipo. Otro ejemplo usualmente conocido se encuentra en las encuestas de hogares, donde cada onda se le pregunta al individuo cual es su estado civil. Si reporta estar casado, siendo que en la onda anterior reportó soltero, no se le pide la fecha exacta de su casamiento. La información disponible dice que entre una onda y otra el individuo cambio de estado marital. Bajo este último concepto, puede suceder que exista gran cantidad de observaciones que

experimentaron el evento durante el mismo intervalo de tiempo. Este problema se denomina “observaciones empatadas”.

C) Empates

Los empates se producen cuando existen observaciones repetidas en un mismo momento del tiempo. Empíricamente, la medida de tiempo en la cual se registra el dato de duración es generalmente en años, trimestres o meses. En estos casos, es posible encontrar más de una observación con un mismo “tiempo de supervivencia”. Es decir, si bien quizás la finalización de la duración se realiza en momentos distintos, el registro se hace en un determinado instante igual para todos los individuos. Entonces hay que adaptar la metodología a la “posibilidad de existencia de observaciones repetidas”. Existen distintas aproximaciones para el tratamiento de este problema que se analiza en detalle en la bibliografía sobre duración ya citada.

D) Múltiples causas de finalización

Una última cuestión hace referencia a la posibilidad de incorporar en la modelización más de una alternativa en la finalización de un evento determinado, por ejemplo “finalizar los estudios”. Un alumno puede finalizar por haber tomado la decisión de abandonar o bien dicha finalización surge cuando el individuo se gradúa. A este tipo de modelos donde no existe una única causa de finalización se lo denomina “competing risk” y la técnica de duración permite un tratamiento (eficiente) de dicha cuestión.

Siguiendo a Lancaster (1990), el análisis de las funciones asociadas con cada uno de los posibles destinos puede realizarse análogamente como si existiera un único destino, considerando toda la información disponible y tratando a la otra parte de la información como censurada. En este caso particular, cuando se estudia el grupo de alumnos graduados, la información sobre los que desertaron y los que continúan se tomará como censurada. Contrariamente, al estudiar a los desertores, la parte censurada serán tanto los que continúan como los que ya se graduaron.

Es importante aclarar que el análisis aquí realizado define al estudiante como un agente racional que toma sus decisiones maximizando sus preferencias dadas sus restricciones. Bajo este contexto, permanecer en la universidad un tiempo determinado implica una “elección” que realiza el individuo. Esta “estructura de elección introducida en el modelo es lo que diferencia el análisis econométrico de datos de transición del análisis estadístico estándar aplicado a los datos de duración”. Lancaster (1990).

2. Modelización

Este apartado tiene como finalidad presentar algunos conceptos necesarios para analizar los resultados. Primero se presentan las definiciones básicas que se aplican a un grupo de individuos que se suponen “homogéneos”. La segunda parte presenta y desarrolla el modelo principal a estimar y justifica su elección⁹. Como se verá a continuación, más que especificar la forma estructural del modelo, se plantea una “forma reducida”. Se considera también que se está frente a un “modelo de múltiples causas”. Y finalmente, se permite controlar por heterogeneidad no observable.

Con población homogénea

Un primer paso es conocer si la probabilidad de que un evento ocurra en un momento del tiempo (condicionado a que no ocurrió en ningún momento anterior) es constante, creciente o decreciente en

el tiempo. Detectar cuáles son los años de mayores riesgos de deserción, no sólo interesa a los efectos de esta investigación, sino que es útil para el diseño de políticas preventivas de abandono.

Una función de particular importancia que describe este comportamiento es la denominada “función de riesgo”, que representa una secuencia de probabilidades condicionales, $f(t) = \Pr(\text{dejar en el momento } t \text{ dado que estudió hasta } t-1)$. Si se define a T como una variable aleatoria discreta no negativa que representa la *duración de los estudios*, T puede tomar valores $0 \leq t_0 < t_1 < t_2 < t_3 \dots$. Esta variable hará referencia al *tiempo* que un estudiante permanece activo¹⁰ en la universidad e incluirá además información sobre la *forma* de finalización. En el caso estudiado, esta forma de finalización puede adoptar dos destinos diferentes: *graduarse* (obtener el título de grado dentro del período de observación) o *abandonar* (no rendir materia alguna durante el año lectivo o no intentar regularizar). La función de probabilidad asociada a la variable aleatoria discreta T es:

$$f(t_i) = P(T = t_i) \quad i = 0, 1, 2, \dots$$

Si lo que interesa es saber cuál es la probabilidad de que T sea mayor o igual que un valor t , se especifica la denominada función de supervivencia que viene dada por:

$$S(t) = \sum_{j | t_j \geq t} f(t_j)$$

y expresa la probabilidad de que la duración del evento sea $T \geq t$. Si no existiese censura, la estimación de esta función es simplemente contar cuántos llegaron hasta el momento t_j en relación al total.

Como se dijo anteriormente, una función de particular importancia es la función de riesgo¹¹ en t_j que viene definida como:

$$I_j = P(T = t_j | T \geq t_j) = \frac{f(t_j)}{S(t_j)} \quad j = 0, 1, 2, \dots$$

y expresa la probabilidad de que un individuo finalice en t_j siendo que “sobrevivió” hasta t_j .

Luego, puede escribirse: $1 - I_j = 1 - \frac{f(t_j)}{S(t_j)}$ y de aquí se demuestra que la función de supervivencia es

igual a: $S(t_j) = \prod_{i=0}^{j-1} (1 - I_i)$. Una estimación no paramétrica de la función de supervivencia sería entonces:

$$\hat{S}(t) = \prod_{j | t_j < t} (1 - \hat{I}_j).$$

Si existe censura, para ciertos individuos se observa sólo que su duración llegó “por lo menos hasta un momento determinado”, pero no la duración completa. Entonces, la estimación de \hat{I}_j es: del total de individuos que estuvieron en el momento j (conjunto de riesgo n_j), a cuántos le ocurre el evento en ese momento j (h_j). Es decir, $\hat{I}_j = \frac{h_j}{n_j}$. Este estimador es conocido como el estimador

no paramétrico de Kaplan-Meier. Obsérvese que este estimador es una función únicamente de los datos de la muestra. El uso de este estimador permite describir la relación existente entre la función de riesgo y el tiempo de duración hasta que ocurre el evento. Este estimador puede interpretarse como un estimador máximo-verosímil, (Kalbfleisch y Prentice, 1980).

Con población heterogénea

Un segundo paso es investigar si el riesgo I de que un evento ocurra difiere sistemáticamente entre individuos. Es decir, se busca identificar las variables explicativas “de la heterogeneidad observada”¹²

en función del riesgo. Por ejemplo, si el riesgo de desertar disminuye en el tiempo, la pregunta de interés es: ¿qué características distinguen a los estudiantes con “alto” riesgo de desertar de aquellos estudiantes con “bajo” riesgo de desertar?.

La heterogeneidad observada se introduce en el modelo considerando p variables explicativas Z_p siendo $p=1,2,\dots,P$ que caracterizan a los miembros de una población. Por ejemplo, Z_1 representa el género, Z_2 educación de la madre, etc. Se tiene entonces un vector: $\mathbf{Z}_{ij} = [z_{1ij}, z_{2ij}, z_{3ij}, \dots, z_{pij}]$ donde cada elemento del vector representa la característica p para el individuo i en el momento j .

Para incluir el efecto de este vector de características en la duración y riesgo de un evento, las alternativas de modelización más comunes son el Modelo de Falla Acelerada AFTM (Accelerated Failure Time Model) o el Modelo de riesgo proporcional PHM (Proportional Hazard Model).

La diferencia básica entre estas dos alternativas de modelización es la manera de introducir los efectos de las variables explicativas. Es decir, para el caso AFTM, el efecto de las variables explicativas se produce directamente sobre el tiempo de duración, mientras que en el modelo PH el efecto es sobre la función de riesgo.

En el modelo de riesgo proporcional propuesto por Cox (1972) que se desarrolla a continuación, el efecto del vector \mathbf{Z}_{ij} se produce de manera multiplicativa sobre la función de riesgo mediante un factor que no depende del tiempo de duración.

Sea $\mathbf{I}_0(t)$ la función de riesgo para un individuo para el cual el valor de todas las variables explicativas que se encuentran en el vector \mathbf{Z}_{ij} son iguales a “cero”. Esta función $\mathbf{I}_0(t)$ se denomina función de riesgo *base*.

La función de riesgo para el individuo i -ésimo puede escribirse como una proporción de esta función base, es decir:

$\mathbf{I}(t; z) = \mathbf{I}_0(t) \mathbf{y}(z)$, donde $\mathbf{y}(z_i)$ es una función de los valores del vector de variables explicativas¹³ para el i -ésimo individuo. Dicha función se interpreta como el riesgo en el momento t para un individuo cuyo vector de variables explicativas es \mathbf{Z}_i , relativo al riesgo para un individuo cuyo \mathbf{Z} es cero.

Como el riesgo relativo $\mathbf{y}(z_i)$ no puede ser negativo, es conveniente suponer una función $\mathbf{y}(z_i) = e^{(\mathbf{m})}$, donde $\mathbf{m}_i = b_1 z_{1i} + b_2 z_{2i} + \dots + b_p z_{pi}$. El \mathbf{m} es llamado componente lineal¹⁴ para el individuo i -ésimo. El modelo de riesgo proporcional general viene a ser entonces:

$$\mathbf{I}_i(t/z) = e^{(b_1 z_{1i} + b_2 z_{2i} + \dots + b_p z_{pi})} \mathbf{I}_0(t)$$

La función de riesgo base es “la misma para todos los individuos”¹⁵. Por lo tanto, para dos individuos con variables explicativas iguales a \mathbf{Z} y \mathbf{Z}^* , el cociente de sus funciones de riesgo es:

$$\frac{\mathbf{I}(t/z)}{\mathbf{I}(t/z^*)} = \frac{e^{\left[\sum_{p=1}^p b_p z_p\right]} \mathbf{I}_0(t)}{e^{\left[\sum_{p=1}^p b_p z_p^*\right]} \mathbf{I}_0(t)} = e^{\left[\sum_{p=1}^p b_p (z_p - z_p^*)\right]}$$

Esta razón (denominada riesgo relativo) es constante en el tiempo; luego, las tasas de riesgo son proporcionales. La interpretación de los coeficientes esta dada por la siguiente derivada:

$$\frac{\partial \ln \mathbf{I}(t/z)}{\partial z_p} = \mathbf{b}_p$$

con $p = 1, 2, \dots, P$. \mathbf{b}_p da el cambio proporcional en la función de riesgo que resulta de un cambio marginal en la p -ésima variable explicativa (que está en “niveles”). Es decir, los coeficientes tienen la interpretación de semi-elasticidades de la función de riesgo con respecto a las variables explicativas.

Si ahora \mathbf{Z} y \mathbf{Z}^* sólo difieren en la p -ésima variable explicativa, la cual es una variable binaria, se tiene que:

$$\frac{\mathbf{I}(t/z)}{\mathbf{I}(t/z^*)} = e^{\mathbf{b}_p}$$

En este modelo, hay tres supuestos implícitos. Un primer supuesto hace referencia a la inexistencia heterogeneidad no observada, ya que toda la heterogeneidad presente entre los individuos es recogida en las variables explicativas. Un segundo supuesto de proporcionalidad y un último de linealidad.

Las ventajas de este modelo PH (modelos de tiempo de supervivencia) sobre AFM (modelos de duración acelerada) son varias. En primer lugar, la interpretación del PH es más sencilla e indica el efecto de la variable explicativa sobre el riesgo de que ocurra un evento (ej. abandonar o graduarse). En segundo lugar, tanto el tratamiento para incorporar variables explicativas cambiantes en el tiempo como el planteo de “múltiples razones de finalización” son más fáciles bajo el modelo de PH. Una última ventaja, y quizás la más importante, es la posibilidad de estimar los efectos de las variables explicativas sobre el riesgo sin necesidad de especificar una función paramétrica para el riesgo base.

Si se supone que los datos son generados por un modelo de riesgo proporcional de tiempo continuo pero observados sólo en “intervalos de tiempo discretos” $(t_i - 1, t_i]$, o bien alternativamente se supone que la duración es intrínsecamente discreta, Prentice y Gloeckler (1978) han demostrado que la correspondiente función de riesgo en tiempo discreto viene dada por:

$$\mathbf{I}_j(z_{ij}) = 1 - \exp\left\{-e^{[b_1 z_{1i} + b_2 z_{2i} + \dots + b_p z_{pi}] + \mathbf{I}(t)}\right\}$$

donde $\mathbf{I}_j(z_{ij})$ son las tasas de riesgo en tiempo discreto para la persona i en cada intervalo de duración de $j = 1, \dots, t$. La diferencia fundamental con el modelo continuo es la interpretación de la función de riesgo. En el caso discreto es la probabilidad condicional mientras que en el caso continuo hace referencia a la tasa instantánea de riesgo. Cada elemento del vector \mathbf{Z} representa una característica para el individuo i en el momento j , y el vector de coeficientes \mathbf{b} (parámetros a estimar) es idéntico al vector \mathbf{b} del modelo de riesgos proporcionales presentado inicialmente. Finalmente, $\mathbf{I}(t)$ es la función (de riesgo base) que describe la duración y puede ser estimada en forma no paramétrica¹⁶.

Esta última modelización¹⁷ es aplicada en este trabajo. Esta elección se basa en dos fundamentos principales.

Por un lado, se considera “conceptualmente” más correcto, dado el tipo de datos con los que se cuenta. Además, para el evento considerado (especialmente, la deserción), no se conoce el momento exacto en que se realiza la decisión, sólo se sabe que ocurre “dentro de un determinado intervalo de tiempo”.

Por otro lado, el modelo discreto tiene varias ventajas en la “interpretación” y “verificación” de los supuestos. Al permitir incorporar las variables dicotómicas asociadas a los distintos momentos del tiempo (o sea los parámetros de la función de riesgo base), la modelización discreta proporciona una estimación directa de dicha función base a partir de la cual se puede construir la función de supervivencia base y medir las duraciones medias para diferentes subconjuntos de individuos en función de sus características socio-económicas. Otra ventaja es que se puede verificar de manera sencilla el supuesto de proporcionalidad del modelo. Es decir, verificar si los efectos de las variaciones en las características de los individuos sobre el riesgo son “independientes” del momento del tiempo en que se mida. Si se introducen las interacciones entre las variables dicotómicas del tiempo y la variable explicativa y éstas son significativas, luego no se puede asegurar que se cumple el supuesto de proporcionalidad ya que el efecto es diferente en cada momento del tiempo. Nótese que en este caso, el supuesto de proporcionalidad es menos restrictivo que para el caso continuo, ya que los efectos sólo deben ser proporcionales en los intervalos y no en cada instante del tiempo.

Con respecto al supuesto de la inexistencia de heterogeneidad no observable, Meyer (1990), propone introducir en el modelo la posibilidad de existencia de heterogeneidad no observable entre individuos de la siguiente manera:

$$I_j(z_{ij}) = 1 - \exp \left\{ -e^{[b_1 z_{1i} + b_2 z_{2i} + \dots + b_p z_{pi}] + I(t) + \log(e)} \right\}$$

La diferencia de este modelo de Meyer con el anterior de Prentice y Gloeckler es la incorporación de un nuevo término que resume la heterogeneidad no observable, representada por e : una variable aleatoria con distribución Gamma¹⁸ (con media uno y $Var = S^2$). Esta variable aleatoria resume el impacto de un conjunto de factores que afectan al riesgo de que ocurra el evento pero no son observados en forma directa, ya sea por ser “intrínsecamente inobservables” o porque no están disponibles en los datos. O alternativamente puede interpretarse en términos de “errores de medición” en los datos.

Debido a la “potencial fragilidad de los modelos que incorporan heterogeneidad no observable” Jenkins (1995), se estimaran “ambos” modelos. Los modelos son estimados mediante el método de máximo-verosimilitud y técnicas no paramétricas para la obtención de la función de riesgo base¹⁹.

Definiendo un indicador de censura $c_i = 1$ si la duración de la persona i -ésima se observa completamente y $c_i = 0$ si la duración esta censurada, la función de log-verosimilitud para el segundo modelo es:

$$\sum_{i=1}^N \log [(1 - c_i) A_i + (1 - c_i) B_i], \text{ donde :}$$

$$A_i = \left[1 + \sum_{j=1}^{t_i} \{ \exp(I_{ij} + \ln(Var)) \} \right]^{-\frac{1}{Var}}$$

$$B_i = \left[1 + \sum_{j=1}^{t_i-1} \{ \exp(I_{ij} + \ln(Var)) \} \right]^{-\frac{1}{Var}} - A_i, \quad \text{si } t_i > 1$$

$$B_i = 1 - A_i, \text{ si } t_i = 1, \text{ donde:}$$

$$I_{ij} = [b_1 z_{1i} + b_2 z_{2i} + \dots + b_p z_{pi}] + I(t)$$

La función log-verosimilitud para el primer modelo²⁰ es el caso límite cuando la $Var \rightarrow 0$.

Previo a la estimación de estos modelos, es necesario realizar modificaciones en la estructura de la base de datos, que se describen en el **Anexo B** “Reordenamiento de la base de datos”.

IV. DATOS

El modelo de la sección 2 capítulo III puede ser aplicado a cualquier muestra de estudiantes. En esta investigación se utilizaron los datos de la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario. En principio, esta elección se debió fundamentalmente a la disponibilidad de información que brindó la Universidad.

Se tuvo acceso a los datos sobre las características socioeconómicas de los alumnos que son recabadas en los formularios del Sistema Unificado de Registración (SUR). Estos formularios deben ser completados por los ingresantes a la universidad y actualizados anualmente en la reinscripción, con los cambios que se hubieran verificado²¹.

También se contó con los datos sobre las materias rendidas (aprobadas o no) y regularizadas, con nota obtenida y fecha de presentación a los exámenes²². Estos datos son de interés para poder realizar el seguimiento académico de los estudiantes y detectar su condición de desertor, graduado o alumno que continua sus estudios.

Se trabajó con la cohorte de *ingresantes de 1991* a la carrera de Contador Público de la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario.

Una cohorte se define en este trabajo como el conjunto de estudiantes que se inscribe a una carrera universitaria en un año determinado. Tiene la ventaja, a diferencia de trabajar con todo el grupo de alumnos existentes en un momento determinado del tiempo, de permitir realizar el seguimiento de un conjunto de estudiantes que se encuentran en las mismas condiciones iniciales y que enfrentan las mismas circunstancias académicas, sociales y otras.

Para realizar la elección del año a estudiar se tuvo en cuenta la cohorte que disponía de mayor información y además se trató de que la duración de los estudios tenga una variabilidad suficiente de manera que fuera posible realizar un análisis estadístico que captara las diferencias existentes en las trayectorias académicas de los alumnos, de modo que cuanto más alejada en el tiempo se encuentre la cohorte, mejor.

Según la tabla de inscripciones a la Facultad, el total de personas que se anotó para comenzar la carrera llegó a 1423. De este grupo, hay 23 alumnos que no completaron sus fichas de ingreso ni tampoco intentaron cursar ni rendir ninguna materia, por lo que no se los consideró en el conjunto a estudiar. Para los 1400 individuos restantes se realizó un seguimiento académico desde marzo de 1991 a mayo de 2001. Debido a que de este total, 24 estudiantes tienen matrícula cancelada ya sea por mal comportamiento, o por haberlos encontrado en falta grave (por ejemplo copiando en un examen), también se excluyó del grupo bajo estudio, quedando así un conjunto de *1376 estudiantes*.

En la *Tabla 1* se observa la distribución de frecuencia conjunta de la situación académica y la duración de los estudios para la cohorte de 1991. Se computa, para cada año, la cantidad de alumnos que abandonan, continúan o se gradúan. Aquí caben dos aclaraciones importantes.

En primer lugar, cuando se hace referencia al “*j-ésimo año*”, este incluye t_j pero excluye t_{j-1} , es decir $(t_{j-1}, t_j]$. Por ejemplo, el año 2 del cuadro abarca desde el 31 de marzo de 1992 (t_{j-1} intervalo abierto) al 31 de marzo de 1993 inclusive (t_j intervalo cerrado), y así sucesivamente para los demás años. El último año (11) abarca desde el 31 de marzo de 2001 (abierto) hasta fines de mayo (última fecha con información disponible al momento de recolectar los datos académicos).

En segundo lugar, se define al desertor como aquel estudiante que durante un período lectivo de un año completo no intentó ni regularizar algún parcial ni rendir materias, independientemente de aprobar o no. Es decir, el alumno se encontró “pasivo” durante ese año sin observársele actividad académica alguna. En algunas investigaciones a este comportamiento se lo suele llamar “primera deserción” (first drop-out) ya que no se puede saber si luego del período de observación el individuo retomará o no sus estudios o bien decidirá iniciar otra carrera..

Los estudiantes agrupados en “continúan” son aquellos que en mayo de 2001 aún se encuentran activos en la Facultad y no han tomado la decisión de desertar o egresar. Los graduados son todos aquellos alumnos que, cumpliendo los requisitos académicos, han aprobado el total de materias y han obtenido su título de grado. Como puede observarse en la *Tabla 1*, este último grupo es muy reducido. Del total de 1376 ingresantes en 1991, luego de algo más de 10 años, sólo lograron recibirse 240 alumnos (17,44%); 920 estudiantes (66.86%) desertaron y 216 (15.70%) aun continúan sus estudios²³.

Tabla 1
Distribucion de frecuencia conjunta de la situacion academica del alumno y duraci3n de los estudios

		Período de observación (Años*)										Total	
		Duración teórica					Años restantes de observación						
		1	2	3	4	5	6	7	8	9	10		11
		0	0	0	0	0	0	0	0	0	29		187
Situación Académica	Continúan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.11	13.59	15.70
		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	13.43	86.57	100.00
		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	40.28	95.90	15.70
		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	40.28	95.90	15.70
	Abandonan	298	177	107	81	77	42	39	40	41	12	6	920
		21.66	12.86	7.78	5.89	5.60	3.05	2.83	2.91	2.98	0.87	0.44	66.86
		32.39	19.24	11.63	8.80	8.37	4.57	4.24	4.35	4.46	1.30	0.65	100.00
		100.00	100.00	100.00	100.00	84.62	50.60	41.49	42.55	48.81	16.67	3.08	66.86
	Se gradúan	14	41	55	54	43	31	2	240				
		0.00	0.00	0.00	0.00	1.02	2.98	4.00	3.92	3.13	2.25	0.15	17.44
		0.00	0.00	0.00	0.00	5.83	17.08	22.92	22.50	17.92	12.92	0.83	100.00
		0.00	0.00	0.00	0.00	15.38	49.40	58.51	57.45	51.19	43.06	1.03	17.44
	Total	298	177	107	81	91	83	94	94	84	72	195	1376
		21.66	12.86	7.78	5.89	6.61	6.03	6.83	6.83	6.10	5.23	14.17	100.00

*Años se refiere al calendario academico.

El año 11 abarca desde 31 marzo 2001 a fines mayo 2001 (última informaci3n disponible)

Fuente: Elaboraci3n propia en base al SUR I - FCEyE UNR

Cada celda de este cuadro contiene la frecuencia absoluta, la frecuencia relativa, el porcentaje con respecto a la fila y el porcentaje con respecto a la columna, desagregado según situaci3n académica y año de observaci3n²⁴.

De la observaci3n de la tabla surgen algunos resultados interesantes:

1. La mayor deserci3n se da en los primeros años. Como indica la tercera fila del grupo de alumnos que abandonaron, un 32.39 % de este grupo desertó durante el primer año. Este porcentaje desciende a 19.24 % en segundo año y vuelve a bajar en los años restantes.
2. Los mayores porcentajes de graduados se encuentran en el séptimo y octavo año. Se evidencia entonces una “demora” en la finalizaci3n de los estudios según el plan teórico de cinco años. La permanencia en la Universidad un mayor tiempo a lo te3ricamente requerido se experimenta para casi el total de egresados. Sólo el 5.83% de los graduados finaliza en tiempos curriculares.

En el sexto año, el porcentaje de estudiantes que se gradúa es aproximadamente igual al porcentaje de alumnos que abandonan, ambos con respecto al total de ingresantes.

Las características socioeconómicas disponibles para cada estudiante se refieren a: variables indicativas de la educación y orientación previa recibida por el alumno (como ser tipo de colegio secundario al que asistió, título obtenido, si asistió previamente a otra carrera universitaria); variables demográficas y personales del estudiante (género, edad, estado civil, tipo y lugar de residencia); variables indicativas de su situación laboral y variables relacionadas con las características de los padres de los ingresantes (categoría ocupacional y educación de sus padres).

Una descripción detallada cada una de estas variables incluídas en el modelo y sus estadísticas resúmenes se presentan en el **Anexo A** “Descripción de las variables”. Todas las variables presentadas, si bien pueden cambiar a lo largo del tiempo, se consideran constantes durante los años y se trabaja con los valores declarados por los individuos “al iniciar” la universidad (exceptuando la situación laboral del alumno que corresponde a la última vez que se reinscribió en la facultad).

Por último, cabe señalar que la base de datos originariamente construída incluye un registro por persona. Para realizar la estimación de los modelos planteados, es necesario construir previamente una “base orientada a periodos”. Es decir, incluir un registro para cada período de tiempo en el cual el individuo está en riesgo de experimentar el evento²⁵.

V. RESULTADOS EMPÍRICOS

Se estiman dos modelos diferentes para cada uno de los evento de interés (deserción y graduación). Un primer modelo (**Modelo 1**) de riesgo discreto propuesto por Prentice y Gloeckler (1978). Y un segundo modelo propuesto por Meyer (1990), que incluye la heterogeneidad no observable entre individuos (**Modelo 2**)²⁶. Es decir, se modela también la existencia de un termino de error aleatorio, que viene a representar todo aquello que afecta el riesgo de abandonar o la probabilidad de recibirse pero que no es observado por el econometrista.

Ambos modelos utilizan el total de 1376 estudiantes²⁷. Dado que se tiene un evento con “múltiples causas de finalización” mutuamente excluyentes, se estiman por separado como si existiese una única causa de finalización. Por un lado, se analiza al grupo de desertores considerando a los graduados y a los que continúan como censurados. Por el otro, al estudiar los factores relacionados con la mayor probabilidad condicional de graduación, los datos de los desertores y los que continúan son los que se encuentran censurados.

1. Factores relacionados con el riesgo de deserción

Los resultados de las estimaciones se reportan en la *Tabla 2*. Cuando el coeficiente estimado es positivo, indica que la variable afecta positivamente a la probabilidad condicional de desertar. Expresando en riesgo relativo e^b indica cuánto más riesgoso es el grupo con respecto a la categoría base, dado todo lo demás constante. El “efecto” de una variable sobre la función de riesgo es proporcional y no cambia en el tiempo.

El valor p del test de razón de verosimilitud es cero, indicando que el modelo con heterogeneidad no observable es estadísticamente significativo. Debe tenerse en cuenta que este test no es una regla de decisión entre un **Modelo 1** y **Modelo 2**²⁸, ya que sólo dice si la heterogeneidad no observable es significativamente distinta de cero.

Si bien los coeficientes para el modelo 2 son mayores en magnitud, la dirección de los efectos no varía de un modelo a otro y los coeficientes significativos en un modelo también lo son en el otro. El hecho que los coeficientes estimados sean mayores en valor absoluto puede deberse a que al no tener en cuenta la heterogeneidad no observada, induce a una sobre-estimación de la tasa de riesgo y esto atenúa la magnitud del impacto de las variables explicativas. (Jenkins, 2000).

El primer grupo de variables explicativas se refiere a la educación y orientación previa que tuvo el individuo antes de ingresar a realizar sus estudios universitarios. Controlando por heterogeneidad no observable (**Modelo 2**), las variables relacionadas con la clase de colegio secundario al que asistió el estudiante son significativas para explicar las diferencias en el riesgo de deserción. Los resultados indican que los egresados de colegios nacionales tienen cuatro veces más probabilidad de desertar que los egresados de colegios dependientes de la universidad, dado todo lo demás constante. Esto puede estar reflejando una diferencia en la calidad de educación previa recibida. Una menor preparación en el nivel secundario, hace que el estudiante tenga que realizar un esfuerzo (individual) mucho mayor durante los primeros tiempos en la universidad.

Tabla 2 Factores relacionados con el riesgo de deserción

Variables	Modelo 1			Modelo 2		
	Coef.	St. Err.	ExpB	Coef.	St. Err.	ExpB
Magist. Otros	0.023	0.110	1.024	0.398	0.312	1.488
Bach. Nac.	-0.002	0.099	0.998	0.086	0.256	1.090
Col. Nac.	0.488 ***	0.159	1.628	1.392 ***	0.368	4.025
Col. Prov. Mun.	0.665 ***	0.173	1.946	1.571 ***	0.402	4.813
Col. Priv. Relig.	0.503 ***	0.158	1.654	1.203 ***	0.343	3.331
Col. Priv. Part.	0.421 **	0.180	1.524	1.240 ***	0.399	3.457
Col. Otros	0.686 **	0.318	1.987	1.591 **	0.772	4.909
Egreso Sec	-0.783 ***	0.102	0.457	-2.092 ***	0.360	0.123
Cont. otra carr.	0.276	0.189	1.318	0.673	0.576	1.961
Term. otra carr.	-0.054	0.209	0.947	1.869 *	1.015	6.482
Aband. otra carr.	-0.238 *	0.129	0.788	-0.327	0.382	0.721
O. V. Priv	0.184	0.123	1.203	0.351	0.280	1.420
O. V. Ofic.	0.168	0.152	1.184	0.243	0.343	1.275
O. V. Prof.	-0.162	0.150	0.850	-0.282	0.331	0.754
Edad	0.192 ***	0.063	1.212	0.695 ***	0.215	2.005
Edad2	-0.002 *	0.001	0.998	-0.008 **	0.004	0.991
Varon	0.146 **	0.069	1.158	0.314 *	0.163	1.369
Soltero	0.524 ***	0.125	1.690	1.340 ***	0.344	3.822
Reside Flia	0.23 ***	0.087	1.264	0.728 ***	0.217	2.071
Rosarino	0.345 ***	0.088	1.413	0.753 ***	0.221	2.124
Sit. Lab. Ingr.	0.454 ***	0.101	1.576	1.230 ***	0.283	3.422
Sit. Lab. Sal.	-0.501 ***	0.103	0.605	-1.418 ***	0.304	0.242
Fallec. Desc.	-0.034	0.167	0.966	-0.070	0.467	0.932
Capatz-Encarg.	-0.264 *	0.139	0.767	-0.159	0.322	0.853
Jefe-Direc.-A.J	-0.254 **	0.108	0.775	-0.733 ***	0.256	0.480
Dueño	-0.113	0.107	0.893	-0.179	0.243	0.836
Cuenta Propia	-0.051	0.092	0.950	-0.521 **	0.229	0.594
Ed. P. Baja	0.3785 **	0.189	1.460	1.000 *	0.514	2.721
Ed. P. Media	0.4113 ***	0.125	1.509	0.708 **	0.277	2.032
Ed. P. Media A.	0.2541 **	0.124	1.289	0.745 ***	0.290	2.108
Ed. M. Baja	0.4609 ***	0.177	1.586	1.053 **	0.453	2.868
Ed. M. Media	0.198	0.129	1.220	0.464 *	0.279	1.591
Ed. M. Media A	0.197	0.126	1.218	0.405	0.265	1.500
Gamma var exp(ln_varg)				2.944 ***	0.611	
Log Likelik.	-2435			-2378		
LR test (1) vs (2)				113.31		
Pr. Test Chi(1)				0.000		
Nro obser.	7281			7281		

*** Significativo al 1%

** Significativo al 5%

* Significativo al 10%

Esto, en muchos casos, puede inducir al alumno a abandonar los estudios superiores. Es importante destacar que los colegios dependientes de la Universidad son gratuitos pero no libres, sino que se requiere un exámen de ingreso que es selectivo.

El haber concluido el secundario e iniciado seguidamente la universidad reduce la probabilidad condicional de desertar con relación a quien esperó algún tiempo entre la finalización de sus estudios secundarios y el inicio de sus estudios superiores.

Algunos estudios Spady (1970) Tinto (1975) concluyen que quienes ya asistieron alguna vez a la universidad poseen menos riesgo de desertar con respecto al grupo que inicia por primera vez estudios universitarios (principiantes), simplemente porque el primer grupo ya “se ambientó” al nivel universitario. Sin embargo, según los resultados aquí obtenidos, fijando un nivel de confianza del 5%, no existen diferencias significativas en el riesgo de abandonar entre quienes iniciaron y abandonaron ya otra carrera y los principiantes. Las mismas conclusiones se obtienen para quien continúa simultáneamente o terminaron otra carrera, dado todo lo demás constante.

Las variables que indican si el ingresante recibió una orientación vocacional previa al inicio de la carrera tampoco son significativas para explicar la probabilidad condicional de deserción²⁹. El abandono por “fracaso vocacional” no parece ser un factor importante que explique el riesgo de deserción.

Igualmente, dado todo lo demás constante, un alumno que obtuvo un título secundario de bachiller nacional con respecto a otro alumno que tuvo un título comercial (y comenzó la carrera de Contador Público) no difieren significativamente en el riesgo de deserción.

Los resultados obtenidos indican que las características demográficas y personales del estudiante son todas significativas al 1% (excepto varón significativa al 10%) para explicar las diferencias en las probabilidades condicionales de abandono. Ser varón influye positivamente en incrementar el riesgo de deserción; se estima que el riesgo del varón es 1.36 veces mayor que el de la mujer. La literatura también señala este resultado, atribuyéndolo a factores culturales o bien a que el varón es menos perseverante al continuar una carrera que la mujer. Ocurre algo similar para los solteros, con respecto a quienes están casados, viudos o separados. Aquí pueden estar interactuando dos efectos contrapuestos. Por un lado, se esperaría que quien esté soltero tenga menos responsabilidades familiares y más tiempo dedicado al estudio, controlando por todas las demás variables. Este efecto reduciría el riesgo de desertar para los solteros. Sin embargo, existe también el efecto del “costo de oportunidad del tiempo”, supuestamente menor para los solteros que para los casados, que juega incrementando el riesgo de abandono. Este último efecto sería el que domina.

Los estudiantes que viven con sus familias tienen mayor riesgo de desertar que los alumnos que tienen que vivir en forma independiente, manteniendo todo lo demás constante. El mismo resultado se observa para los alumnos de Rosario, en relación con estudiantes que vienen de otra ciudad o pueblo. Los que provienen de otras localidades tienen “costos hundidos” (alquiler, mudanza, adaptación) de iniciar la carrera que son virtualmente nulos para quienes ya son de la ciudad.

Por último, interesa indagar si los ingresantes de mayor edad son más propensos a abandonar que los estudiantes más jóvenes, dada la posible mayor restricción de tiempo que enfrentan los primeros. Efectivamente, el cambio proporcional en la función de riesgo que resulta de un aumento en un año en la *edad* del individuo que ingresa a la Universidad, es positivo. La variable edad al cuadrado capta la relación no lineal en forma de U invertida entre el riesgo y la edad.

Como era de esperarse, haber iniciado los estudios universitarios y simultáneamente estar trabajando provoca un mayor riesgo de abandono (3.4 veces más) con relación a quienes inician sus estudios sin una responsabilidad laboral. Estos últimos estudiantes se supone que deben asignar su tiempo total entre el estudio o el ocio. Sin embargo, dado todo lo demás constante, trabajar “el último año”, disminuye la probabilidad condicional de deserción con relación a un alumno que no trabaja.

El último grupo de variables relacionadas con las características de los padres reflejan parte del entorno familiar en que se encuentra el estudiante. Para el conjunto de variables indicativas de la última ocupación que desarrolló el padre, la categoría base elegida es “obrero o empleado”. Los signos de los coeficientes que acompañan a estas variables son, en todos los casos, negativos (aunque no todos significativos). Si el estudiante tiene padre que es jefe, director o alto jefe su riesgo de abandonar los estudios es 26.7% menor con relación a un individuo con padre obrero o empleado. Un caso similar sucede con los estudiantes con padres que son cuentapropistas. Ambos coeficientes, son significativos fijando un nivel de significancia del 1% y 5%, respectivamente.

Otro punto de interés es analizar cuan importante es la educación de los padres en la decisión de no concluir los estudios superiores. Según los resultados de este estudio, cuanto menor nivel educativo de los padres mayor riesgo de deserción. Una explicación es que cuanto más educación formal hayan incorporado los padres, más valor le otorguen a más años de estudio. Un alumno con padre que sólo tiene educación primaria incompleta o no tiene educación formal, tiene un riesgo de desertar que es 2.7 veces mayor que el de la categoría base (padres con estudios superiores completos). Todos los coeficientes de este grupo (excepto educación media alta de la madre) tienen un efecto sobre el riesgo de deserción (imponiendo un nivel de significancia del 5% y 10%)³⁰.

Resultados similares se obtienen para el total de estudiantes de la Universidad Nacional de Rosario, aplicando otras técnicas. En Perfil Social (1990), la identificación de grupos de mayor riesgo de abandono señalaba a los alumnos que trabajan y tienen padres obreros con sólo nivel primario y también a estudiantes que viven en Rosario, trabajan y tienen padres empleados que alcanzaron sólo educación primaria. Tanto la educación que poseen los padres de los alumnos, como las ocupaciones en donde trabajan, son factores “exógenos” al estudiante, que difícilmente pueda cambiar el alumno. Es decir, los riesgos se relacionan más con cuestiones estructurales. Sería de interés estudiar mecanismos alternativos que permitan a estudiantes con ciertas características desfavorables reducir sus riesgos de deserción.

Una pregunta que se planteaba al inicio de este estudio era saber *cuándo* era más probable que ocurriría la deserción. Es importante conocer si la probabilidad condicional de desertar es constante en el tiempo o si existen períodos en donde los riesgos se incrementan. En la siguiente tabla se presenta la estimación “no paramétrica” de los coeficientes de la función de riesgo base para los modelos 1 y 2 analizados anteriormente. Las variables binarias creadas para cada intervalo diagnostican la “forma” de la función (que no se impuso a priori). La probabilidad de abandonar los estudios durante un intervalo determinado, dado que el alumno llegó hasta ese momento, varía a lo largo de la trayectoria académica. Valores más grandes los coeficientes (menos negativos), se asocian con riesgos más grandes. Según el **Modelo 1**, durante los primeros años se experimenta el mayor riesgo. En el **Modelo 2**, el riesgo se incrementa año a año hasta el año quinto³¹.

Tabla 3 Función de riesgo base

	Modelo 1			Modelo 2		
	Coef.	Std. Err.	P> z	Coef.	Std. Err.	P> z
Año 1	-5.735	0.879	0.000	-14.373	2.996	0.000
Año 2	-5.838	0.877	0.000	-13.348	2.913	0.000
Año 3	-6.105	0.880	0.000	-13.045	2.869	0.000
Año 4	-6.215	0.880	0.000	-12.778	2.839	0.000
Año 5	-6.125	0.879	0.000	-12.352	2.814	0.000
Año 6	-6.590	0.885	0.000	-12.542	2.797	0.000
Año 7	-6.508	0.885	0.000	-12.238	2.782	0.000
Año 8	-6.274	0.885	0.000	-11.741	2.765	0.000
Año 9	-6.000	0.885	0.000	-11.122	2.745	0.000
Año 10	-6.989	0.919	0.000	-11.861	2.742	0.000
Año 11	-7.377	0.965	0.000	-12.140	2.753	0.000

2 Factores relacionados con la graduación

La duración teórica de la carrera universitaria bajo análisis es de cinco años. Sin embargo, los alumnos que no desertan prolongan sus estudios por un mayor tiempo. Esta sección se concentra en una pregunta: ¿qué factores “favorecen” a la finalización exitosa de la carrera universitaria dado el tiempo que lleva el alumno estudiando?. Los trabajos de Cameron y Heckman (1998), y Cameron y Tabor (1999) aplicando otras técnicas sugieren que factores tales como el “entorno familiar” juegan un papel central en la determinación de las decisiones de finalizar los niveles educativos iniciados. Los resultados encontrados en este trabajo apoyan esta evidencia. Según los modelos estimados para el grupo de graduados³², el nivel de educación alcanzado por sus padres influye significativamente en la probabilidad de graduarse. Un estudiante cuyo padre sólo cuenta con primario incompleto (o sin instrucción) tiene una probabilidad de graduarse 70% menor que un alumno con padre profesional. Esta brecha se reduce cuanto mayor educación haya adquirido el padre. Para las demás categorías educativas y observando el **Modelo 1**, los coeficientes son todos significativos al 1% y al 5% (excepto educación del padre media alta que es significativa al 10% y educación de la madre baja, que no es significativa).

A diferencia del caso de deserción, los resultados ahora sugieren que la heterogeneidad no observable no es significativa. Se prefiere entonces realizar el análisis de los coeficientes para el **Modelo 1** (suponiendo que no existe heterogeneidad no observable).

Dado todo lo demás constante, para un estudiante que está trabajando (al menos) durante el último año en que se asiste a la universidad, la probabilidad de graduarse es 45% menor que para un alumno con tiempo completo dedicado al estudio. En cambio, el efecto de iniciar la carrera trabajando no es significativo en reducir la probabilidad de graduarse.

Tabla 4 Factores relacionados con la probabilidad condicional de graduarse

Variables	Modelo 1			Modelo 2		
	Coef.	St. Err.	ExpB	Coef.	St. Err.	ExpB
Magist. Otros	-0.063	0.259	0.938	-0.054	0.294	0.947
Bach. Nac.	-0.387 *	0.221	0.679	-0.436 *	0.261	0.647
Col. Nac.	-0.657 ***	0.230	0.518	-0.699 **	0.273	0.497
Col. Prov. Mun.	-0.601 **	0.299	0.548	-0.626 *	0.348	0.535
Col. Priv. Relig.	-0.670 ***	0.222	0.511	-0.719 ***	0.268	0.487
Col. Priv. Part.	-0.074	0.248	0.929	0.002	0.326	1.003
Col. Otros	0.069	0.764	1.072	-0.064	0.942	0.938
Egreso Sec	-0.242	0.329	0.785	-0.339	0.421	0.712
Cont. otra carr.	0.715	0.594	2.045	0.835	0.756	2.306
Term. otra carr.	1.496 **	0.654	4.465	1.518 *	0.807	4.565
Aband. otra carr.	-1.403 **	0.578	0.246	-1.590 **	0.707	0.204
O. V. Priv	-0.203	0.250	0.816	-0.197	0.292	0.821
O. V. Ofic.	-0.074	0.320	0.928	-0.068	0.373	0.934
O. V. Prof.	0.512 **	0.228	1.670	0.631 **	0.322	1.880
Edad	0.306	0.509	1.358	0.289	0.601	1.336
Edad2	-0.009	0.012	0.991	-0.009	0.014	0.990
Varon	-0.347 **	0.144	0.706	-0.392 **	0.182	0.675
Soltero	0.616 **	0.266	1.852	0.729 **	0.357	2.074
Reside Flia	-0.271	0.167	0.762	-0.277	0.195	0.758
Rosarino	-0.310 *	0.170	0.733	-0.329	0.204	0.719
Sit. Lab. Ingr.	0.087	0.171	1.091	0.040	0.210	1.041
Sit. Lab. Sal.	-0.585 ***	0.165	0.557	-0.632 ***	0.205	0.531
Fallec. Desc.	0.222	0.444	1.249	0.243	0.503	1.275
Capatz-Encarg.	-0.264	0.356	0.767	-0.287	0.400	0.750
Jefe-Direc.-A.J	-0.005	0.203	0.995	-0.058	0.257	0.943
Dueño	0.136	0.209	1.146	0.192	0.259	1.213
Cuenta Propia	0.122	0.187	1.130	0.133	0.218	1.143
Ed. P. Baja	-1.192 **	0.538	0.303	-1.287 **	0.614	0.276
Ed. P. Media	-0.528 ***	0.194	0.590	-0.6323 **	0.282	0.531
Ed. P. Media A.	-0.353 *	0.182	0.703	-0.417 **	0.237	0.659
Ed. M. Baja	-0.078	0.437	0.924	-0.084	0.499	0.919
Ed. M. Media	-0.638 ***	0.199	0.528	-0.666 ***	0.242	0.514
Ed. M. Media A	-0.501 ***	0.180	0.606	-0.504 **	0.217	0.604
Gamma var exp(ln_varg)				0.5621	0.99	
Log Likelik.	-750.73			-749.87		
LR test (1) vs (2)				0.41		
Pr. Test Chi(1)				0.52		
Nro obser.	3132			3132		

*** Significativo al 1%

** Significativo al 5%

* Significativo al 10%

Por otro lado, existe una mayor probabilidad de recibirse para el caso de las mujeres y también para los estudiantes solteros. Sin embargo, ingresar a la facultad con más edad o bien residir con la familia durante el ciclo lectivo, no explican las diferencias en las chances de concluir la carrera. Tampoco haber recibido orientación vocacional por una entidad oficial o privada (con respecto a quien no recibió orientación alguna) parece tener efecto sobre la probabilidad condicional de recibirse. Si, en

cambio, la orientación vocacional fue dada por un profesional, el alumno tiene 1.67 veces más chance de culminar exitosamente su carrera.

Los coeficientes que acompañan a las variables binarias indicativas del colegio secundario al que asistió el ingresante son significativas al 1% y 5% y tienen signo negativo revelando que el haber asistido a un colegio dependiente de la universidad favorece la probabilidad de graduarse con relación a cada uno de los demás colegios secundarios. Específicamente, para un egresado de un colegio nacional la probabilidad condicional de graduarse es 48,2% menor que para un alumno egresado de un colegio secundario dependiente de la universidad.

Egresar del secundario e iniciar al año siguiente la carrera de grado no parece contribuir significativamente a una mayor probabilidad condicional de culminar exitosamente los estudios universitarios-sin embargo, sí constituía un factor importante al momento de explicar las diferencias en las tasas de riesgos de deserción.

Finalmente, haber culminado otra carrera superior (antes de iniciar esta carrera) hace que el alumno, dado todo lo demás constante, tenga 4.46 veces más probabilidades de recibirse que un estudiante que inicia por primera vez una carrera universitaria. Quien abandonó otra carrera, en cambio, tiene una probabilidad de obtener el título universitario 75.4% inferior con respecto al principiante.

Los modelos estimados supusieron tiempo discreto. Cabe aclarar que cuanto menor es el tiempo entre sucesivas “reexaminaciones” (por ejemplo, en vez de observar cada año quiénes abandonaron, realizarlo semestralmente) y mayor es el número de intervalos, más se aproximarían estos resultados a los de un modelo de tiempo continuo.

VI CONCLUSIONES

Durante las últimas décadas, la universidad pública argentina ha experimentado un incremento en la cantidad de alumnos inscriptos, no así en el número de graduados. Incluso, del reducido grupo que completa sus estudios universitarios, la mayoría extiende su permanencia en la Universidad un tiempo mayor al teóricamente establecido. Tanto la deserción como la demora en la graduación en las universidades públicas son fenómenos claros y socialmente conocidos.

Este trabajo analiza estas cuestiones e investiga cuáles son los factores que se relacionan con la probabilidad de desertar o graduarse que tiene un estudiante, *condicionado* al tiempo que lleva en la facultad, enfocando de esta manera el proceso de finalización de los estudios desde una perspectiva dinámica.

Se utiliza un modelo de riesgo proporcional discreto propuesto por Prentice y Gloeckler (1976). Sin especificar ninguna forma funcional a priori de la función de riesgo base, se estima también el modelo propuesto por Meyer (1990). Este modelo, además de la incorporación de las variables explicativas observables, permite introducir heterogeneidad “no observable” entre individuos.

La aplicación empírica se realiza para la cohorte de estudiantes que ingresaron a la carrera de Contador Público en la Universidad Nacional de Rosario en 1991. Una vez creada la base de datos, se efectúa el seguimiento año a año de esta cohorte hasta mayo de 2001. Al finalizar la recolección de datos existen alumnos de este grupo que aun continúan sus estudios universitarios y a los que no se les observa el evento de interés (deserción-graduación); solo se sabe que “estudiaron hasta ese momento”. Este problema de censura en los datos es tratado por la metodología aplicada en este trabajo.

Los resultados obtenidos sugieren que el factor relacionado con la educación de los padres y el tipo de colegio secundario al que asistió el alumno son importantes en la explicación de las diferencias de riesgo de deserción y probabilidades condicionales de graduación. Independientemente de las demás características, un estudiante cuyo padre cuenta con primario incompleto (o sin instrucción) tiene un 70% menos posibilidades de graduarse que aquel estudiante con padre profesional. Controlando por heterogeneidad no observable, el riesgo de desertar es 2.86 veces mayor para el alumno cuya madre cuenta con primaria incompleta con respecto a un estudiante con madre con estudios superiores completos. También se comprobó que el riesgo de deserción es significativamente menor (26.7% menor) para alumnos cuyos padres son jefes, directores o altos jefes en comparación con estudiantes de padres obreros o empleados. A su vez, dado lo demás igual, iniciar la carrera y estar trabajando hace que el estudiante tenga 3.4 veces más riesgos de abandonar. El género también juega un papel significativo. Se estima que el riesgo de abandono para el varón es 1.36 veces mayor que el de una mujer. Las demás características demográficas y personales (residir en Rosario, vivir con la familia, ser soltero y tener más edad al iniciar la carrera) resultaron también significativas afectando positivamente la probabilidad condicional de abandonar. Egresar del secundario e iniciar seguidamente la universidad resultó ser un factor relevante para el caso de deserción, pero no ha sido un factor que incremente significativamente la probabilidad de graduarse que tiene un estudiante. Por último, el estudiante que ya abandono otra carrera en comparación con aquel que inicia por primera vez la facultad tienen menos chances de graduarse. Controlando por las demás variables, el alumno que se encontraba trabajando la última vez que se reinscribió a la universidad tiene menos probabilidad de finalizar exitosamente la carrera que quien no trabajaba en ese momento.

Si bien conclusiones obtenidas en este trabajo presentan sólo una primera aproximación a los fenómenos de deserción y graduación y los factores influyentes, son útiles tanto a nivel facultad - para implementar estrategias de ingreso y seguimiento de los estudiantes, y para las regulaciones de la condición de alumno regular -, como también a nivel de diseño de políticas educativas que permitan mejorar el sistema educativo universitarios - por ej. las vinculadas con el financiamiento de las universidades y de los estudiantes . Los factores que cumplen un rol importante en incrementar las probabilidades condicionales de graduación y en disminuir el riesgo de deserción están íntimamente relacionadas con el entorno familiar y deben ser un tema de preocupación para la política social y su relación con la política universitaria propiamente dicha.

Existen varias cuestiones a considerar en futuros trabajos. Si se almacenara la información sobre los cambios de la situación socioeconómica de los alumnos, sería interesante incluir el efecto de estas variables explicativas “cambiantes” en el tiempo. Adecuar el sistema informático para no perder esta información es una tarea pendiente. También podría introducirse al modelo alguna medida de rendimiento de modo de investigar, por ej. porque dos alumnos con igual rendimiento toman decisiones diferentes de finalización de sus estudios. De contar con los datos para toda la universidad habría posibilidades de captar también diferencias por carreras.

ANEXOS

A. Descripción de las variables

En los modelos a estimar, se incluyen como regresores (**Z**) todas aquellas variables disponibles que afectan el riesgo de deserción y la graduación. A estas variables se las divide en **4 grupos**:

1. Variables indicativas de la educación y orientación previa recibida por el alumno

Título secundario obtenido. Son tres categorías posibles: Magist. Otros=1 si obtuvo título magisterio, técnico agropecuario u otros (0 en otro caso); Bach. Nac.=1 si obtuvo título de bachiller nacional (0 en otro caso); Comercial=1 si es perito mercantil o comercial (0 en otro caso) (escogida como categoría base).

Clase colegio al que asistió el alumno. Son seis categorías posibles: Col. Nac.=1 si fue a un colegio secundario nacional; Col. Prov. Mun.=1 si fue a un colegio Provincial o Municipal; Col. Dep. Univ.=1 si asistió a un colegio dependiente de la Universidad (elegida como categoría base); Col. Priv. Relig.=1 si concurrió a un colegio privado religioso; Col. Priv. Part.=1 si privado particular y Col. Otros=1 si fue a “otros” colegios (ej. militares).

Orientación Vocacional recibida previo a ingresar a la universidad. Tiene cuatro categorías posibles: Sin O.V.=1 si no recibió ninguna orientación vocacional (es categoría base); O.V. Priv=1 si recibió orientación por una entidad privada; O. V. Ofic.=1 orientación otorgada por una entidad oficial y O.V. Prof.=1 si recibió orientación por un profesional especializado.

Esta variable cuenta con cuatro categorías. La categoría base es Empieza=1 si es la primera vez que el alumno comienza un estudio superior. Cont. otra carr.=1 si además de la carrera iniciada, continua otra carrera. Term. otra carr.=1 en caso en que el estudiante, al iniciar esta carrera, ya haya terminado previamente algún otro estudio superior. Y por último, Aband. otra carr.=1 en caso que abandono una carrera anteriormente³³.

Egreso Sec. : esta variable toma valor 1 si el estudiante egreso del secundario e inicio la universidad y 0 en otro caso.

2. Variables demográficas y personales del estudiante

Edad: Años de edad al momento de ingresar la Universidad

Varón: variable binaria que toma valor 1 si el estudiante es varón, cero en otro caso

Soltero: variable binaria que toma valor 1 si el alumno es soltero, cero si es casado viudo o separado

Reside Flia: Tipo de residencia, toma valor 1 si el estudiante reside con su familia, valor 0 si reside en forma independiente, o en residencia universitaria.

Rosarino: Lugar de procedencia del estudiante, toma valor 1 si es de Rosario y 0 en otro caso

3. Variables indicativas de la situación laboral del alumno

Sit. Lab. Ingr.: Variable binaria que toma valor 1 si el estudiante trabajaba “al iniciar” la carrera, es decir, en el momento que realiza su primera inscripción.

Sit. Lab. Sal.: Variable binaria que toma valor 1 si el alumno trabaja “el último año que estuvo (se reinscribió) en la Universidad”.

Si bien lo optimo seria poder contar con información año a año del estado laboral del individuo, por las restricciones de datos ya comentadas se cuenta sólo con estos dos momentos del tiempo.

4. Variables relacionadas con características de sus padres

Categoría ocupacional del padre del alumno. Incluye seis categorías (de 10 posibles): Fallec. Desc.=1 si el padre es desconocido o fallecido; Capataz Encarg.=1 si es artesano y técnico, capataz o encargado; Jefe-Direc.-A.J.=1 si es Jefe, Gerente, director o alto jefe o componente de sociedades u organizaciones con fines de lucro; Dueño=1 si es dueño; Cuenta Propia=1 Independiente, cuenta propia sin personal; Obrero=1 si el padre es obrero o empleado (categoría base).

Máximo nivel de educación alcanzado por el padre del estudiante. Se las agrupo en cuatro categorías: Ed. P. Baja= nivel de educación del padre baja (no hizo estudios o escuela primaria incompleta); Ed. P. Media=1 nivel de educación del padre media baja (escuela primaria completa o secundaria incompleta); Ed. P. Media A.= nivel de educación del padre medio alto (colegio secundario completo o estudios universitarios o superiores incompletos); Ed. P. Alta= nivel de educación del padre alta (estudios universitarios o superiores completos).

Máximo nivel de educación alcanzado por la madre del alumno (M). (*Idem* anterior.).

Este conjunto de variable son reflejo de los posibles ingresos de las familias de los estudiantes. Se espera que cuanto mayor educación posean los padres, podrán ocupar trabajos mas remunerados y por consiguiente obtener mejores ingresos.

En el siguiente cuadro se resumen las estadísticas descriptivas de estas variables.

Estadísticos descriptivos de las variables

	Total Alumnos				
	Obs	Mean	Std. Dev.	Min	Max
Edad	1376	19.331	3.745	16	50
Bach. Nac.	1376	0.105	0.306	0	1
Col. Nac.	1376	0.148	0.355	0	1
Comercial	1376	0.748	0.434	0	1
Col. Nac.	1376	0.318	0.466	0	1
Col. Prov. Mun.	1376	0.148	0.355	0	1
Dep. Univ.	1376	0.078	0.268	0	1
Col. Priv. Relig.	1376	0.335	0.472	0	1
Col. Priv. Part.	1376	0.108	0.311	0	1
Col. Otros	1376	0.013	0.114	0	1
Egreso Sec	1376	0.709	0.454	0	1
Empieza	1376	0.868	0.338	0	1
Cont. otra carr.	1376	0.029	0.168	0	1
Term. otra carr.	1376	0.025	0.155	0	1
Aband. otra carr.	1376	0.078	0.268	0	1
Sin O.V.	1376	0.793	0.405	0	1
O. V. Priv	1376	0.090	0.286	0	1
O. V. Ofic.	1376	0.054	0.226	0	1
O. V. Prof.	1376	0.063	0.243	0	1
Varon	1376	0.466	0.499	0	1
Soltero	1376	0.891	0.312	0	1
Reside Flia	1376	0.775	0.417	0	1
Rosarino	1376	0.792	0.406	0	1
Sit. Lab. Ingr.	1376	0.488	0.500	0	1
Sit. Lab. Sal.	1376	0.645	0.479	0	1
Fallec. Desc.	1376	0.042	0.201	0	1
Capatz-Encarg.	1376	0.070	0.256	0	1
Jefe-Direc.-A.J	1376	0.175	0.380	0	1
Dueño	1376	0.164	0.370	0	1
Cuenta Propia	1376	0.236	0.425	0	1
Obrero	1376	0.313	0.464	0	1
Ed. P. Baja	1376	0.057	0.231	0	1
Ed. P. Media	1376	0.496	0.500	0	1
Ed. P. Media A.	1376	0.304	0.460	0	1
Ed. P. Alta	1376	0.143	0.350	0	1
Ed. M. Baja	1376	0.070	0.256	0	1
Ed. M. Media	1376	0.474	0.499	0	1
Ed. M. Media A	1376	0.333	0.471	0	1
Ed. P. Alta	1376	0.123	0.328	0	1

B. Reordenamiento de la base de datos

La base de datos armada para el total de estos estudiantes es una “base orientada a persona”, es decir, cada individuo tiene una fila de datos. El individuo *i-esimo* tiene información sobre:

- ✓ Duración: Longitud de tiempo (o intervalos) en el que el individuo estuvo activo en la universidad. Es decir, la longitud de tiempo en el cual el estudiante estuvo expuesto al riesgo de experimentar el evento.
- ✓ Indicador de Censura: Indica si el estudiante experimento el evento de interés durante el período en el cual se lo observo ($c = 1$); o bien si se encuentra censurado ($c = 0$).
- ✓ Variables explicativas: para los factores que no varían en el tiempo, existe un único valor registrado para cada individuo.
- ✓ A los fines de realizar la estimación, esta base de datos debe ser transformada³⁴ en una “base orientada a periodos”. La reorganización implica que cada individuo posee ahora tantas filas como

periodos de observación registrados para dicho individuo. Siguiendo la definición utilizada, significa que existirán t_i filas para cada persona $i = 1, \dots, N$. Luego, la nueva base de datos cuenta con:

- ✓ Indicadores del tiempo: Conjunto de variables binarias $D_{1ij}, D_{2ij}, \dots, D_{Jij}$.
- ✓ Variables explicativa
- ✓ Indicador del evento: Una variable Y que toma el valor 1 si ocurrió el evento y 0 en otro caso. El siguiente cuadro ilustra la base original y la transformada para el ejemplo de que el evento de interés sea la deserción.

Base de datos original orientada a persona

ID	Duracion	Censor	Soltero	Varon
1	3	1	1	0
2	1	1	0	1
3	9	0	1	0
4	2	1	1	1

Base de datos transformada orientada a periodos

ID	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	Soltero	Varon	Y
1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
1	0	1	0	0	0	0	0	0	0	0	0	1	0	0
1	0	0	1	0	0	0	0	0	0	0	0	1	0	1
2	1	0	0	0	0	0	0	0	0	0	0	0	1	1
3	1	0	0	0	0	0	0	0	0	0	0	1	0	0
3	0	1	0	0	0	0	0	0	0	0	0	1	0	0
3	0	0	1	0	0	0	0	0	0	0	0	1	0	0
3	0	0	0	1	0	0	0	0	0	0	0	1	0	0
3	0	0	0	0	1	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	1	0	0	0	0	0	1	0	0
3	0	0	0	0	0	0	1	0	0	0	0	1	0	0
3	0	0	0	0	0	0	0	1	0	0	0	1	0	0
3	0	0	0	0	0	0	0	0	1	0	0	1	0	0
4	1	0	0	0	0	0	0	0	0	0	0	1	1	0
4	0	1	0	0	0	0	0	0	0	0	0	1	1	1

La primera columna identifica a cada individuo (ID).

En este caso existen 11 intervalos en los que se puede observar la duración, el individuo 1 “sobrevivió sólo los tres primeros intervalos” y dentro del tercer intervalo le ocurrió el evento (deserto). Por esto, la columna Y en la tercera fila correspondiente al individuo 1 toma valor 1. El individuo 3, en cambio, no deserta y en el período 9 sale de observación ya que se graduó.

Esta es la forma que tiene que tener la base de datos para poder estimar los modelos en tiempo discreto planteados en este trabajo.

BIBLIOGRAFÍA

- Adelman, C. (1999). *Answers in the Tool Box: Academic Intesity, Attendace Patterns, and Bachelor's Degree Attainment*. Jessup, MD. Department of Education
- Alemanya Leira, R. (1993) "Modelització de la Durada dels Estudis Universitaris: Una Aplicació a la Universitat de Barcelona". Tesis Doctoral no publicada.
- Alemanya Leira, R., Conduras A. y Costa, A. (1990). "Rendiment Academic i permanencia a la Universitat de Barcelona". Consell Social. Universitat de Barcelona.
- Allison, P. D. (1982). "Discrete-time methods for the analysis of event histories". *Sociological Methodology*. Jossey- Leinhardt S. (ed.) Bass Publishers, San Francisco, pp. 61-98.
- Araoz, A. (1968). "Comentarios sobre el trabajo del Dr. Julio Olivera: La Universidad como unidad de producción.". *Revista Económica Año XIV- Numero 1-2 Enero-Agosto 1968*. Universidad Nacional de La Plata.
- Bank, J. B, Slavings, R. L y Biddle, B.J. (1990). "Effects of Peer, Faculty and Parental Influences on Students' Persistence". *Sociology of Education*. Vol. 63, Issue 3. pp.208-225.
- Bean, J. P (1980). "Student attrition, intensions and confidence". *Research in Higher Education* 17. pp 291-320.
- Cabrera, A. F., Nora. A. y Castaneda, M. B. (1993). "College persistence: structural equations modeling test of integrated model of student retention". *Journal of Higher Education* 64(2) pp. 123-139.
- Cerconi, L. y Ortiz de Guevara, E. (1999). "Los estudios universitarios como determinantes del nivel de ingreso: Una aproximación econométrica". III Encuentro Internacional de Encomia. Universidad Nacional de Córdoba-CIEC. Agosto.
- Clotfelter C. T, Ehrenberg, R.G y Getz M; Siegfried, J.J. (1991) *Economic Challenges in Higher Education*. The University of Chicago Press.
- Cox D. R (1970). *The analysis of binary data*. London: Methuen.
- Cox D. R y Oakes D. (1984). *Analysis of Survival data*. London: Chapman and Hall.
- DesJardins, S. L., McCall, B.P., Ahlburg, D.A. y Moye, M. (2001). "Adding a Timing Light to the Toolbox". *Frothcoming in Research in Higher Education*. Vol. 43.
- Gray, M.W y Taylor A. (1989). "A study of factor influencing student performance in mathematics on the Florida college level academic skills tests". *Journal of Negro Educational*. Vol 58.
- Jenkins S. P. (1995). "Easy estimation methods for discrete time duration models". *Oxford Bulletin of Economics and Statistics* 57(1): 129-138.
- Jenkins S. P. (2000). *Introduction to the analysis of spell duration data. Lesson 8: Unobserved heterogeneity ('frailty')*. Institute For Social And Economic Research University Of Essex.

- Kalbfleisch, J. D. y Prentice, R. L. (1980). *The statistical analysis of failure time data*. John Wiley and Sons.
- Kiefer, N. (1988). "Economic duration data and hazard functions". *Journal of Economic Literature*. 26 pp. 646-679.
- Klein, J. y Moeschberger, M. (1997). *Survival Analysis of Failure Data*. New York: Wiley.
- Lancaster, T. (1990). *The econometric analysis of transition data*. (1ra ed.). Cambridge University Press.
- Meyer, B. D. (1990). "Unemployment insurance and unemployment spells". *Econometría* 58(4): pp. 757-782.
- Montoya Diaz, M. D. (1999). "Extended stay at university: an application of multinomial logit and duration models". *Applied Economics* 31 pp. 1411-1422.
- Moscoloni N., Conti O., Tuttolomondo I. y Meinardi, B. (1990). "Perfil Social de los Estudiantes de la Universidad Nacional de Rosario. Año 1990". Documento de Trabajo. Instituto Rosario de Investigaciones en Ciencias de la Educación.
- Pagura J.A; Quaglino M. B. e Iturbide, D. (2000) "Un modelo estadístico para evaluar tiempos medios empleados en culminar etapas en la universidad". *Revista IRICE* pp.129-141.
- Porto, A. y Di Gresia L. (2001). "Rendimiento de Estudiantes Universitarios y sus determinantes" presentado en la Asociación Argentina de Economía Política (AAEP), Noviembre.
- Pascarella, E. T y Terenzini, P.T (1983). "Predicting voluntary freshman year persistence withdrawal behavior in a residential university A path analytic validation of Tinto's model". *Journal of Educational Psychology* 75 pp. 215-226.
- Pascarella, E. T y Terenzini, P.T (1991). *How college affects students: finding and insights from twenty years of research*. Jossey-Bass. San Francisco.
- Prentice R. L. y Gloeckler (1978). "Regression analysis of grouped survival data with application to breast cancer data". *Biometrics* 34 pp.57-67
- Robinson, R. (1990). "Understanding the gap between entry and exit: a cohort analysis of african american students persistence". *Journal of Negro Educational*. Vol. 59.
- Singer y Willet (1991). "From whether to when: New methods for studying student dropout and teacher attrition". *Review of Educational Research* 61 (4), pp. 407-450.
- Sosa Escudero, W. (1999). *Tópicos en econometría aplicada. Notas de clase. Trabajo Docente N° 2*. Departamento de Economía. Universidad Nacional de la Plata.
- Spady, W. (1970). "Dropouts from higher education: an interdisciplinary review and synthesis". *Intechange* 1. pp.64-85.
- Stata Technichal Bulletin: STB-39 sbe17. "Discrete time proportional hazards regression". Vol 7 pp. 109-121.

Tinto, V (1975). "Dropout from higher education: A theoretical synthesis of recent research." *Review of Educational Research* 45. pp. 89-125

NOTAS

¹ En los Anuarios Estadísticos del Ministerio de Educación disponibles en Internet, se encuentra un detalle de estos datos para el total de las universidades argentinas y para cada universidad en forma separada.

² Permanencia entendida como no abandono.

³³ Los costos sociales reducen el output de la economía si los graduados universitarios son más productivos que aquellos estudiantes que abandonaron la carrera universitaria Pascarella y Terenzini (1991).

⁴ La metodología que estos autores aplican se utiliza, con algunas variantes, en el presente estudio y se tratará en detalle en el próximo apartado.

⁵ Montoya Díaz (1999) estudia también el problema de que aquellos que obtienen el título pero alargan su carrera varios años más de los teóricamente necesarios.

⁶ Los datos surgen de una encuesta realizada por el Departamento de Economía de la Universidad Nacional de La Plata (UNLP) a estudiantes de la Facultad de Ciencias Económicas de la UNLP.

⁷ A esta metodología se la denomina también “análisis de supervivencia” (Survival Analysis) en el campo biomédico; “análisis de fiabilidad” (Reliability or Failure Time Analysis) en el campo industrial; análisis de la historia de un evento (Event History Analysis) en sociología y análisis de duración o transición (Duration or Transition analysis) en el campo económico.

⁸ En este trabajo se considerará que la censura es “independiente”.

⁹ Para un exhaustivo análisis de las derivaciones de los modelos, pueden consultarse en los textos de referencia sugeridos.

¹⁰ Un estudiante “activo” se define como aquella persona que rinde (independientemente de si aprueba o no) o intenta regularizar al menos una materia en el año.

¹¹ Una definición suponiendo tiempo continuo es: $I(t) = \lim_{h \rightarrow 0} \frac{Pr(t \leq T < t+h \mid T \geq t)}{h}$ cuando h tiende a 0; conocida en economía como la inversa de la Razón de Mill.

¹² La heterogeneidad observada se refiere al hecho que es información disponible para el investigador.

¹³ Desde aquí en adelante se supone que las variables explicativas son constantes en el tiempo. Sin embargo, bien podrían introducirse variables explicativas cambiantes en el tiempo sin complicar el análisis, cuando se supone tiempo discreto.

¹⁴ Ver Cox y Oakes (1984) para un análisis exhaustivo de este modelo.

¹⁵ Este modelo puede re-expresarse como: $\log\left(\frac{I_i(t)}{I_0(t)}\right) = (\mathbf{b}_1 z_{1i} + \mathbf{b}_2 z_{2i} + \dots + \mathbf{b}_p z_{pi})$ donde el modelo de riesgo proporcional es similar a un modelo lineal para el logaritmo de la razón de riesgo.

¹⁶ La ventaja principal de no suponer una forma funcional a priori para la función de riesgo base es eliminar la posibilidad de estimar los \mathbf{b} de manera inconsistente por especificación incorrecta de la función base.

¹⁷ También conocida como log-log complementaria: $\log(-\log(1 - I_j)) = \{(\mathbf{b}_1 z_{1i} + \mathbf{b}_2 z_{2i} + \dots + \mathbf{b}_p z_{pi}) + I(t)\}$

¹⁸ La elección de la distribución no es importante cuando la función de riesgo base es estimada en forma no paramétrica. Ver Meyer (1990).

¹⁹ Los modelos discretos tienen la ventaja (sobre los modelos continuos) de ser fácilmente estimables incluso incluyendo variables explicativas cambiantes en el tiempo. Ver Jenkins (1995)

²⁰ La función log-verosímil para el modelo 1 es la misma que la log-verosímil para un modelo lineal general de la familia binomial log-log link complementaria. Ver Allison (1982)

²¹ Es importante aclarar que la reinscripción es obligatoria para poder rendir algún examen durante el año lectivo. En el anexo se adjunta el formulario tipo que completa todo alumno inscripto a la UNR.

²² Una vez cargados los formularios de ingreso, mediante la utilización del programa Access, se combinaron las tablas de la base de datos relacional creada por el sistema informático de la universidad, se extrajeron los individuos de interés y se corrigieron los errores y duplicaciones encontrados, de modo de poder contar con una base lo más pulida posible. El tiempo y el esfuerzo destinado a esta tarea no fue menor. La base queda disponible para futuras investigaciones.

²³ En el grupo inicial de Activos, existían 58 alumnos que si bien estaban reinscriptos en el 2001, su última actividad académica estaba entre el año 3 y 5 (según el caso). Por lo tanto, a este subgrupo se los considero dentro del conjunto de los desertores.

²⁴ Las flechas en color azul muestran como se calcula cada ratio.

²⁵ Ver en el Anexo A “Reordenamiento de la base de datos” para mayor detalle.

²⁶ La estimación se realiza utilizando el programa Stata 6.0. Jenkis programó el comando **pgmhaz** para estimar ambos modelos. El modelo 1 es estimado por MV usando el comando **glm**. El modelo 2 es estimado usando el comando **ml deriv0**, comenzando con un valor inicial de beta estimado en el modelo 1. Ver STB-39 para mas detalles.

²⁷ Nótese que el numero total de observaciones que aparecen en el modelo de deserción es 7281 y 3132 para graduación, ya que en el armado de la base, esto representa la suma para todos los alumnos del total de periodos de observación en riesgo.

²⁸ Ya que no son modelos anidados.

²⁹ La categoría base contra la cual deben analizarse los coeficientes es “no haber recibido ninguna orientación vocacional”

³⁰ En todo este análisis se supuso proporcionalidad. Es decir, que el efecto de un cambio en la variable explicativa sobre el riesgo es independiente del “intervalo” de tiempo en que se mida. Este supuesto es mucho menos restrictivo en el modelo discreto que en el continuo, donde la proporcionalidad se tenia que dar “en cada momento del tiempo”. El supuesto se verifico introduciendo interacciones las variables binarias para cada intervalo con las variables explicativas y no eran significativas.

³¹ De todos modos estos resultados hay que interpretarlos con cautela ya que no se testeó la pendiente de la función. La ventaja principal de la estimación no paramétrica de la función base es evitar obtener estimaciones inconsistentes de los coeficientes por incorrecta especificación de la función de riesgo base.

³² Y considerando la información de los desertores y los que continúan como censurada.

³³ Del grupo que ya abandono otra carrera, aproximadamente el 80% había abandonado la carrera de Analista de Sistemas en la UTN.

³⁴ Esta transformación no sólo es necesaria para asegurar que el programa que estime el modelo genere una contribución “correcta” de cada individuo a la función de verosimilitud, sino también permitiría incluir regresores cambiantes en el tiempo sin problema alguno, si es que se contase con estos datos.